D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi **8**, 129-145 (2007)

WEB MADENCİLİĞİ İLE LOG ANALİZİ
(Araştırma Notu)

Data Mining with Log Analysis
(Research Note)

Abdullah BAYKAL [1]

Özet

Veri madenciliği uygulamalarından biri olan web madenciliği, web verileri üzerinde veri madenciliği fonksiyonlarını yerine getirir ve web içerik madenciliği ile web kullanım madenciliği gibi iki bölüme ayrılır. Web içerik madenciliği; web üzerindeki faydalı bilginin keşif ve analizi ile ilgiliyken, kullanıcı erişim desenlerinin bulunması web içerik madenciliği konusuna girmektedir.

Bu çalışmada web erişim günlükleri kullanılarak saldırı tespiti yapılmaya çalışılmıştır. Bu amaçla yapılan çalışma ile Dicle Üniversitesi web/mail sitesi günlükleri üzerindeki erişim desenleri yorumlanmıştır. Sonuç olarak saldırı amaçlı erişimler bulunmaya çalışılmıştır.

***Anahtar Kelimeler: veri** madenciliği, bilgisayar, program, log analizi*

Abstract

Web mining being one of data mining applications executes the web mining functions on web data and comprises two sections being web content mining and web usage mining. Where Web content mining is related to exploration and analysis of useful data on web, web content mining is interested in discovering user access patterns.

In this study it has been tried to make attack determination by using web access blogs. The study carried out for this purpose has interpreted the access patterns on web/mail site blogs. As a result attack-oriented accesses were tried to be detected.

***Keywords:** data mining, computer, program, log analysis*

1. Saldırı Tesbiti

- Bir sisteme yapılan saldırılar genellikle;

- *Harici ataklar*
- *Bir başkası gibi görünme*
- *İmtiyazı kötüye kullanma*
- *Gizli kullanıcılar*

şeklinde ortaya çıkmaktadırlar. Bu saldırıları etkisiz hale getirmek için engelleme, ele geçirme, caydırma, biçim bozma, bulma ve sayaç atakları gibi teknikler kullanılabilir. [1]

[1] Dr.; D.Ü.Bilgi İşlem Daire Başkanlığı, 21280 Kampüs – Diyarbakır, baykal@dicle.edu.tr

Saldırı tespit metotları ise şu şekildedir;
Anormallik Tabanlı
-Normal ve anormal kullanımlar için tipik desenler tespit edilir ve kullanılır.[5]

İmza Tabanlı
-Önceki atakların ve eşleşen desenlerin imzası modellenir.

Otomatik Kurallar
-Tarihsel bilgi kullanılarak akıllı öğrenim algoritmaları ile normal ve saldırı trafiği modellenir.

Kural Merkezli Politika
-Kuralları uzmanlar tarafından belirlenir.

Saldırı tespiti ile ilgili yaklaşımlar iki kategoriye ayrılmaktadır.

- Kötüye Kullanım Tespiti: Saldırıları tanımak için çok iyi bilinen desenlerden faydalanılır.
- Anormallik Tespiti: Saldırıları tanımak için normal kullanım desenlerinden sapma yapanların bulunması şeklindedir.

Bu yaklaşımların ana problemleri ise şunlardır. Kötüye kullanım tespitinde bilinen saldırı desenleri elle kodlanmak zorunda ve ilk kez yapılan saldırıların tespit edilmesi mümkün olamamaktadır. Anormallik tespitinde ise olaylar arasındaki ilişkilerin yakalanması mümkün olamamaktadır. Saldırı tespiti için bir başka yaklaşım veri madenciliği yaklaşımıdır. Veri madenciliği tabanlı yaklaşımda öğrenim ve tespit ajanları bulunmaktadır. Bu yaklaşım akıllı ajan tabanlı bir yaklaşımdır. Öğrenim ajanları, tespit modelleri ile devamlı eğitilir. Tespit ajanları ise saldırıların tespit için güncellenmiş modeller kullanırlar.[1]

Saldırı tespitinde veri madenciliği kullanımının sebepleri ise şunlardır:

- Denetleme (audit) verisi üzerinde normal ve saldırı etkinlikleri kanıt bırakırlar.
- Veri merkezli bakış açısından bakıldığında saldırı tespiti bir veri analiz işidir.
- İstisna saptanması ve hata/alarm yönetimi gibi başarılı uygulamalarla aynı etki alanı içindedir.

İlgili veri madenciliği algoritmaları ise sınıflama, link analizi ve sıralı analizdir . [3]

Saldırı tespit sistemleri saldırının tespit edildiği noktaya göre daha önce de bahsedildiği gibi iki grupta incelenebilir. Bilgisayar tabanlı saldırı tespit sistemleri, bilgisayar üzerinde çalışan ve bilgisayar üzerindeki aktiviteleri izleyen sistemlerdir. Ağ tabanlı saldırı tespit sistemleri ise ağ üzerindeki kritik noktalara yerleştirilmiş ve algılayıcılar ile ağ trafiğini izleyen sistemlerdir. Veri madenciliği saldırı tespiti için kullanılan tekniklerden biridir.

Özellikle saldırı bilgilerinin analiz edilmesinde kullanılır. Server üzerinde tutulan log dosyaları yapılan bağlantı ve istek bilgilerini tutarlar. Bu dosyalardaki bağlantı bilgilerinden saldırı amaçlı bağlantılar bulunabilir.

Hazırlanan yazılımlar sayesinde servera yapılan bağlantı sıklıklarından veya serverdan istenen dosya tiplerinden bir saldırı olayının olup olmadığı ve saldırıyı yapanın kim olduğu rahatça bulunabilir.

Veri madenciliği tekniklerinden sınıflandırma ve kümeleme saldırganı tanıma maksatlı kullanılabilir. Ayrıca saldırılar bazen istisna durumlar olarak ortaya çıktıklarından istisna saptanması uygulaması olarak ta sonuç elde edilebilir.

Saldırı tespit sistemleri modern güvenlik paketlerinin bir parçasıdır. İmza tabanlı bir tanıma sistemidir. Bu metotlar çeşitli veriler içinden özellikler çıkarmak ve uzmanlar tarafından belirtilen kurallara göre saldırıları bulmak için kullanılır. Bu yöntem sisteme bir atak yapıldığında veritabanına atak bilgilerinin girilmesi şeklinde işler. Yeni bir atak geldiğinde bunun bir atak olduğu başta tanınmaz ancak ikinci sefer atak geldiğinde sistem onu tanır ve ona göre muamele yapar. Bunun için bir saldırı gelmeden o önce saldırı için kestirimde bulunmak amacıyla veri madenciliği ve makine öğrenimi algoritmaları kullanılabilir. Böylece yeni ataklar tanınabilir.

Saldırı tespiti konusunda veri madenciliği tekniklerinden kümelemeyi kullanmak sınıflamayı kullanmaktan daha etkin bir yöntemdir. Sınıflama yapılırken önceden bazı değerlerin uzmanlar tarafından girilmesi gerekmektedir fakat kümeleme yapıldığında sistem kendi kendine bilgilenmekte ve yeni bir durumun saldırı olup olmadığını daha kolay belirleyebilmektedir

2. Anormallik Tespiti

Bir sistemin Aktif olarak sınıflandırılabilmesi için sistemin tespit edilen bir saldırıya gerçek-zamanlı olarak (yada buna yakın) cevap vermesi (güvenlik duvarı kurallarını saldırıya göre düzenlemek yada komut konsolunu saldırı hakkında uyarmak gibi) gerekir. Pasif sistemler genelde aktiviteyi kayıt ederler ve daha sonraki bir tarihte incelenmek üzere saklarlar. Bilgisayar tabanlı sistemler hedeflenen sistemler üzerinde bulunurlar. Ağ tabanlı sistemler ağda, hedef ve saldırgan arasında bir yerde bulunurlar ve akan trafiği saldırı olup olmadığını tespit için dinlerler. Genelde ağ tabanlı sistemler ya DMZ (demilitarized zone)`lerde bulunurlar, ağın güvenlik duvarı ile servis sağlayıcı arasında ya da güvenlik duvarı ile iç ağ arasında ya da bunların herhangi bir kombinasyonunda.

Saldırı tespit sistemleri, başlangıcından beri 'saldırı izleri/işaretleri' (attack signatures) fikrine dayanmaktadır. Yani her saldırının kendini diğer saldırılardan ve normal ağ trafiğinden ayıran bazı izleri vardır. Bu pek çok virüs tarayıcısının dizaynına benzemektedir. Sistem trafiği tarar ve bilinen bir saldırınınkine benzeyen bir işaret gördüğünde neye ayarlanmışsa onu yapar

(sistem yöneticisine çağrı mesajı göndermek, güvenlik duvarı kurallarını güncellemek, konsolu haberdar etmek vs.).

Saldırı tespitte, sık rastlanmayan, tanımlanmayan saldırıları tespit olayı ise 'Anomaly Detection'dır. Anomaly tespit sistemi ile ağ normalde bulunan trafik göz ardı edilir ve normal trafikte bulunmayacak bitler ağ yöneticisinin dikkatine sunulur. Bunun belirli bazı avantajları vardır.

Tam olarak 'Güvenli' bir sistem diye bir şey olmadığını hepimiz biliyoruz. Bugün İnternet`e bağlı olan her makinenin güvenliği yenilgiye uğrayabilir. Bunun olmasını engelleyen şeylerden biri sistemdeki açıkların daha keşfedilmemiş olmasıdır. Fakat onlar ordadır. Peki, yeni bir güvenli açığı bulununca ne olur? Açığı bulan kişi büyük ihtimalle açıktan yararlanmak için bir exploit kodu yazar. Bu kod bir süre için ya arkadaşlarla paylaşılır yada kendisine saklar. Doğal olarak bu kod bir süre sonra güvenlikle ilgilenen toplumun eline geçer ve bir yama hazırlanır. Şimdi, exploit kodu yazılmasıyla bir yama hazırlanana kadar geçen süre içinde Saldırı Tespit Sistemleri ne kadar güvenilirdir? Hiç. Çünkü bir saldırıyı tespit edebilmek için o saldırının ağda hedef sisteme yol alırken neye benzediğini bilmek gerekir.

Ağınızdaki güvenliği görüntülemede alternatif bir yol sunmak için bazı varsayımlar yapalım:

A. Ağınıza erişmek isteyen birisini makinenin kablosunu çekmedikçe alıkoyamazsınız.
B. Güvenliğinizi sağlamak için sınırlı kaynaklarınız var (Hepimizin olduğu gibi).

Bu varsayımlara göre neler yapabiliriz? Tabi ki problemi çözmek için insan gücü ve kaynakları arttırabilir, cluster yapısında güvenlik duvarları satın alıp kurabilir, saldırı tespit sistemleri satın alabilir ve ağdaki tüm makineleri güvenli hale getirmeye çalışabilirsiniz. Fakat gerçekten neyi yapmayı ümit edebilirsiniz?

Yapabileceğiniz en iyi şey sistemlere girişi, saldırganın düşündüğünden daha uzun süre uğraşmasını sağlayacak kadar zorlaştırmak olabilir. İkinci olarak ağınızda güvenlik açığı bulundurabilecek servisleri tespit için bir zayıflık taraması yapabilirsiniz. Ve üçüncü olarak güvenlik açığı içerebilecek olan makinelere erişimi, aktif ya da pasif olarak engelleyecek şekilde önlemler alabilirsiniz.

Bunu nasıl yapabilirsiniz. Tüm ağ trafiğini nasıl sorgulayabilirsiniz, hatta dün yazılmış exploit`leri. Anomaly Detection ile.

Etkili bir Anomaly tespit sistemini herhangi bir linux platformu üzerine basit ücretsiz yazılımlar ve az değişiklikler yaparak kurmak mümkün. Bu araçlar, ipchains/ipfwadm, portsentry, logcheck, gnumeric ve bir e-posta adresi. Sistemin çalışması şöyle:

Her sistemde, ipchains/ipfwadm dinlenmeyen portlara giden trafiği kaydedecek şekilde ayarlı. Eğer bu bir web sunucusu ise ve ssh kullanıyorsanız, ipchains 22/tcp ve 80/tcp haricindeki tüm portlara giden paketleri loglasın. Portsentry yi logcheck`i çalıştıracak şekilde ayarlayın.

portsentry -actp kullanın. Logcheck`i alışılmadık aktivitelerde e-posta adresinize mesaj atacak şekilde ayarlayın. Gnumeric ya da diğer bir spreadsheet programı ile her makinede kötü amaçlı trafiğin kayıtlarını tutun. IP adresi, aktivitenin tarih ve saati, kullanılan portlar (kaynak port dahil) saldırganın IPsinin hostname`i, kontak bilgileri ve IP`nin sahipleri gibi bilgileri tutun.

Bu sistem ile ağınıza giren ve ağınıza ait olmayan her paketi izlemiş olacaksınız. Her paketi. Bir düşünün, saldırganın sistemlerinize girebilmesi için hangi servislerin çalıştığını, kullanılan işletim sistemlerini vs. bilmesi gerekiyor. Onun hızını yavaşlatın, olanları görebilin ve karşılık verebilecek zamanınız olsun.

3. Web kullanım madenciliği ile saldırı tespitinin yapılması

Dicle Üniversitesi sunucuları üzerindeki log dosyalarının incelenmesi için, sur makinası üzerindeki apache web server tarafından oluşturulan access_log ve sendmail posta sunucusu tarafından oluşturulan maillog dosyaları kullanılmıştır.

Aşağıda access log dosyasının bir örneği verilmiştir.

Access log dosya deseni:

- İstemde bulunan IP numarası
- İstem Tarihi ve saati
- İstenilen dosya
- İstem protokolü
- Durum kodu ve Dosya büyüklüğü

85.106.225.211 - - [01/Jan/2006:06:34:04 +0200] "GET /cgi-bin/openwebmail/openwebmail.pl HTTP/1.1" 200 5969

85.106.225.211 - - [01/Jan/2006:06:34:05 +0200] "GET /cgi-bin/openwebmail/openwebmail.pl HTTP/1.1" 200 5969

65.54.188.60 - - [01/Jan/2006:06:34:06 +0200] "GET /data/openwebmail/help/en/tutorial/notes/addhelp.html HTTP/1.0" 200 3668

85.106.225.211 - - [01/Jan/2006:06:34:07 +0200] "GET /data/openwebmail/images/backgrounds/Globe.gif HTTP/1.1" 304 0

85.106.225.211 - - [01/Jan/2006:06:34:08 +0200] "GET /data/openwebmail/images/openwebmail.gif HTTP/1.1" 304 0

85.106.225.211 - - [01/Jan/2006:06:34:16 +0200] "POST /cgi-bin/openwebmail/openwebmail.pl HTTP/1.1" 200 1981

85.106.225.211 - - [01/Jan/2006:06:34:22 +0200] "GET /data/openwebmail/images/backgrounds/Globe.gif HTTP/1.1" 304 0

65.54.188.60 - - [01/Jan/2006:06:38:35 +0200] "GET /bilgiedinme/iletisim.htm HTTP/1.0" 200 5586

65.54.188.60 - - [01/Jan/2006:06:38:43 +0200] "GET /bilgiedinme/bilgi_hakki.htm HTTP/1.0" 200 4308

llogların incelenmesinde aw-loganalyzer ve analog gibi programlar kullanılarak anlamlı sonuçlar elde edilmiştir.

4. Web Server İstatistikleri: [Dicle Üniv.]

Analiz edilen tarih aralığı: Paz,01-Ocak-2006 06:33 / Paz,08-Ocak-2006 07:37 (7.04 gün).

4.1.Genel Özet

(**Git**: İlk Sayfa | Genel Özet | Aylık Rapor | Günlük Özet | Saatlik Özet | Site Tipi Raporu | Organizasyon Raporu | Durum Kodu Raporu | Dosya Boyutu Raporu | Dosya Tipi Raporu | Dizin Raporu | Erişim Raporu)

Parantez içindeki değerlerin temsil ettiği gün sayısı: 7 gün-bitiş: 14-Ocak-2006 11:45.

Başarılı erişimler: 4 478 696 (463 829)
Başarılı erişimler günlük ortalaması : 635 776 (66 261)
Sayfalara yapılan başarılı erişimler: 359 515 (40 052)
Sayfalara yapılan başarılı erişimler günlük ortalaması: 51 035 (5 721)
Başarısız erişimler: 54 986 (4 837)
Yönlendirilen erişimler: 10 189 (462)
Enformasyonal status kodlu erişimler: 47 (12)
Erişilen belirgin dosya sayısı: 87 133 (16 565)
Servis verilen belirgin host sayısı: 36 750 (5 922)
Bozuk KAYIT(LOG) dosyası satırları: 1
Transfer edilen bilgi: 71.27 gigabytes (9.72 gigabytes)
Transfer edilen bilgi günlük ortalaması: 10.12 gigabytes (1.39 gigabytes)

4.2.Aylık Rapor

Herbir ünite (▭) sayfaya 10 000 kez erişimi gösterir (sayfadaki nesnelere erişim dahil).

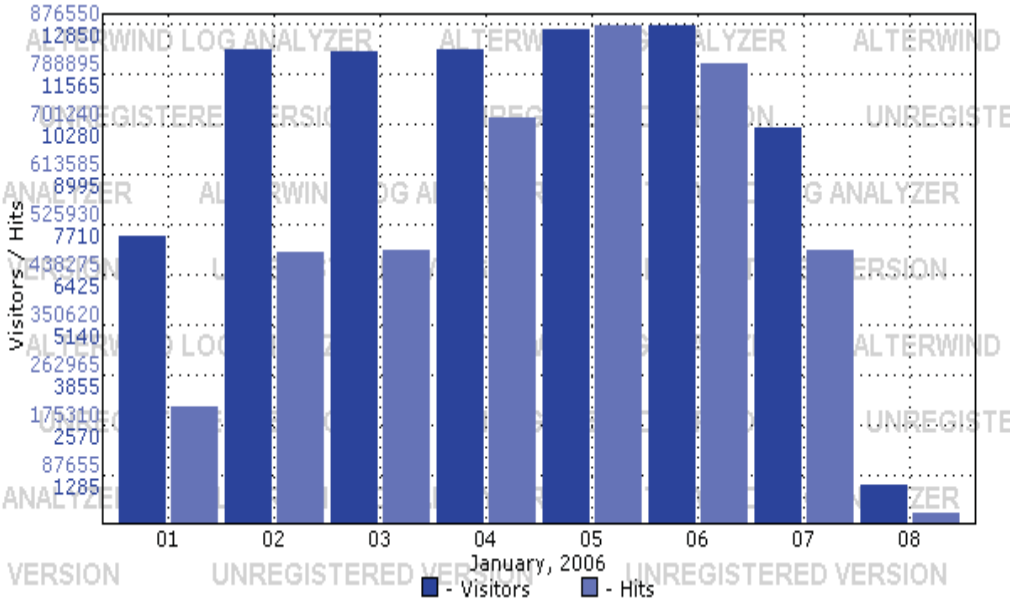
ay	erişim	sayfa	
Ock 006	4478696	359515	

En meşgul ay: Ock 2006 (359 515 kez erişimi gösterir).

4.3.Günlük Özet

Herbir ünite (▭) sayfaya 1 500 kez erişimi gösterir (sayfadaki nesnelere erişim dahil).

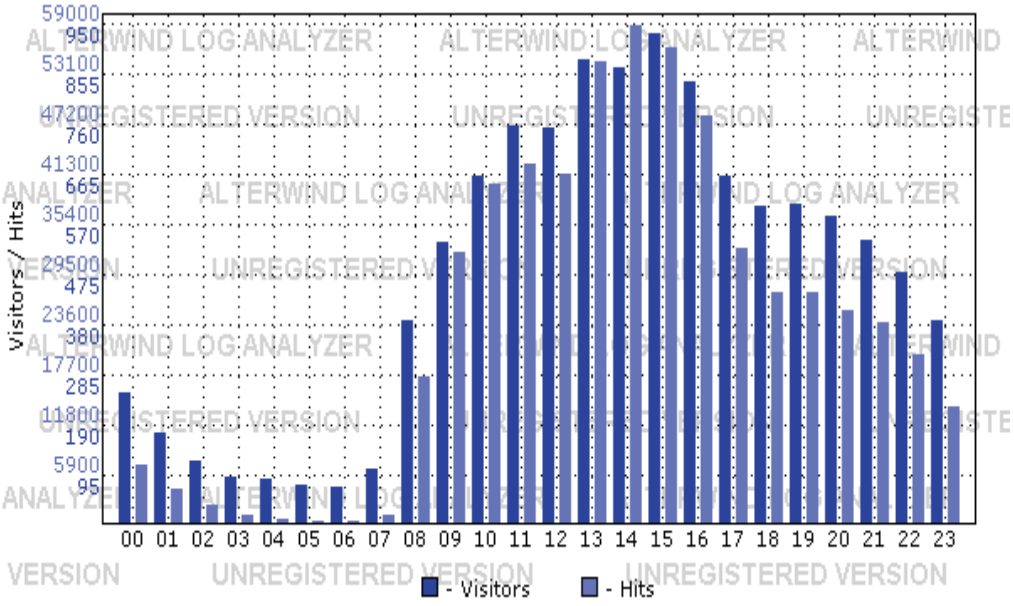
gün	erişim	sayfa
Paz	256891	27355
Pts	532353	50671
Sal	537736	50094
Çrş	763233	64042
Prş	965357	63248
Cum	896708	59491
Cts	526418	44614



4.4.Saatlik Özet

Herbir ünite (▭) sayfaya 800 kez erişimi gösterir (sayfadaki nesnelere erişim dahil).

sa	erişim	sayfa
0	53621	8162
1	31573	5213
2	17965	3941
3	9257	2075
4	5584	1161
5	4634	1470
6	4993	1707
7	11451	2590
8	142014	9997
9	255467	17196
10	318906	21054
11	331299	22247
12	315956	23231
13	419685	28135
14	453944	30538
15	427432	29627
16	367415	26714
17	252227	21601
18	216053	21482
19	208961	15771
20	190081	17155
21	181226	17495
22	152743	16105
23	106209	14848

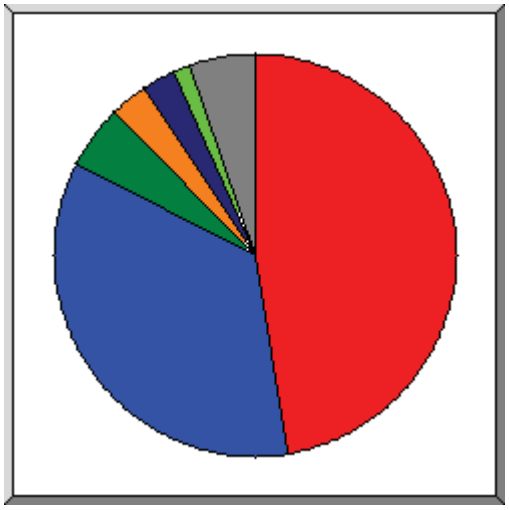


4.5.Site Tipi Raporu

Site tipleri listeleniyor, Sıralama: erişim miktarı.

erişim	bytes%	site tipi
4478696	100%	[Çözümlenmemiş sayısal adres]

4.6. Organizasyon Raporu



Dilimlerin temsil ettiği büyüklük: erişim sayısı.

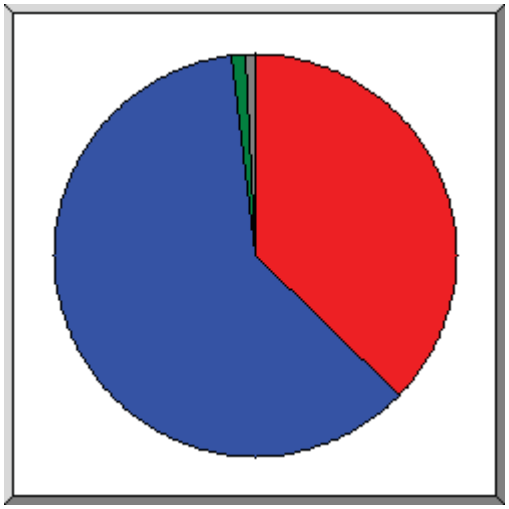
85
10
81.213
81.215
81.214
193.140
diğer

İlk 20 organizasyonlar listeleniyor. Sıralama: erişim sayısı, Sıralama: erişim sayısı.

erişim	bytes%	Organizasyon
2124077	52.96%	85
1569592	19.74%	10
228827	5.61%	81.213
140179	3.87%	81.215
117270	5.34%	81.214
64856	1.20%	193.140
37014	1.72%	212.175
30905	0.48%	70

22272	0.85%	194.27
18161	0.88%	193.255
11672	0.57%	195.175
11322	0.13%	144.122
9464	0.20%	81.212
8042	0.82%	66.249
7165	0.35%	68.142
5102	0.17%	212.174
5099	0.18%	212.156
3793	0.61%	65.54
3766	0.13%	88
3141	0.23%	84
56977	3.98%	[not listed: 1 008 organizasyonlar]

4.7. Durum Kodu Raporu



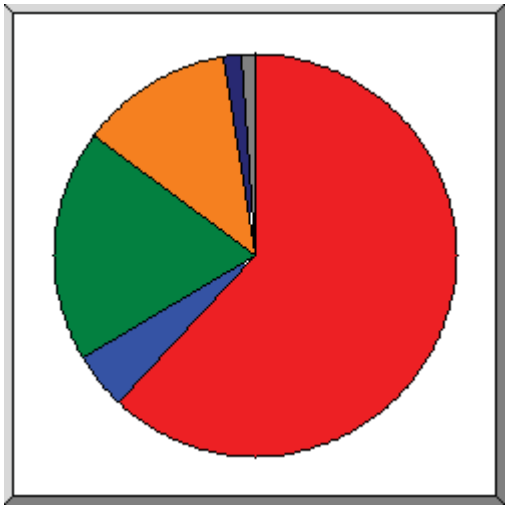
Dilimlerin temsil ettiği büyüklük: erişim sayısı.

200 OK
304 Not modified since last retrieval
404 Document not found
diğer

durum kodları listeleniyor, Numara sıralı.

erişim	durum kodu
47	1xx [Miscellaneous informational]
1694817	200 OK
19301	206 Partial content
4439	301 Document moved permanently
5750	302 Document found elsewhere
2764578	304 Not modified since last retrieval
171	400 Bad request
810	401 Authentication required
155	403 Access forbidden
51266	404 Document not found
1948	405 Method not allowed
628	406 Document not acceptable to client
8	416 Requested range not valid

4.8. Dosya Boyutu Raporu

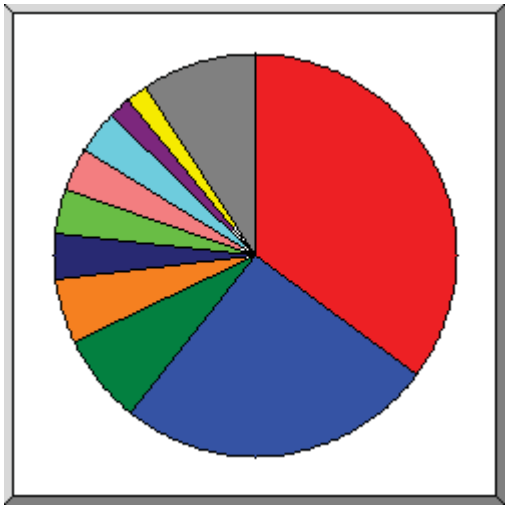


Dilimlerin temsil ettiği büyüklük: erişim sayısı.

0
101B- 1kB
1kB- 10kB
10kB-100kB
100kB- 1MB
diğer

boyut	erişim	bytes%
0	2769970	
1B- 10B	120	
11B- 100B	38333	
101B- 1kB	203341	0.11%
1kB- 10kB	843303	3.04%
10kB-100kB	550793	20.86%
100kB- 1MB	67908	20.36%
1MB- 10MB	4380	14.06%
10MB-100MB	543	39.82%
100MB- 1GB	5	1.75%

4.9. Dosya Tipi Raporu



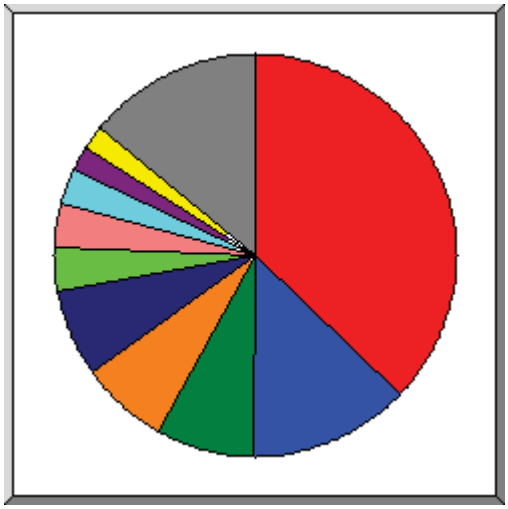
Dilimlerin temsil ettiği büyüklük: erişim miktarı.

.wmv
.jpg [JPEG graphics]
[Dizinler]
.pdf [Adobe Portable Document Format]
.doc [Microsoft Word document]
.pl [Perl scripts]
.htm [Hypertext Markup Language]
.gif [GIF graphics]
.exe [Executables]
.nrg
diğer

dosya tipleri listeleniyor - en az 0.1% trafiği olan, Sıralama: erişim miktarı.

erişim	bytes%	dosya uzantısı
507	35.21%	.wmv
1668493	25.32%	.jpg [JPEG graphics]
228245	7.23%	[Dizinler]
5445	5.31%	.pdf [Adobe Portable Document Format]
6716	3.61%	.doc [Microsoft Word document]
72888	3.57%	.pl [Perl scripts]
66483	3.57%	.htm [Hypertext Markup Language]
2235643	3.44%	.gif [GIF graphics]
510	1.85%	.exe [Executables]
7	1.75%	.nrg
687	1.37%	.ppt
1348	0.97%	.zip [Zip archives]
29531	0.96%	.php [PHP]
64787	0.88%	.html [Hypertext Markup Language]
10382	0.85%	.swf
705	0.72%	.pps
5264	0.66%	.png [PNG graphics]
189	0.57%	.mp3 [MP3 sound files]
55	0.51%	.mpg [MPEG movie]
470	0.36%	.bmp
1049	0.34%	.xls
46	0.14%	.dat
94	0.14%	.mpeg [MPEG movie]
79152	0.67%	[not listed: 72 dosya uzantıları]

4.10. Dizin Raporu



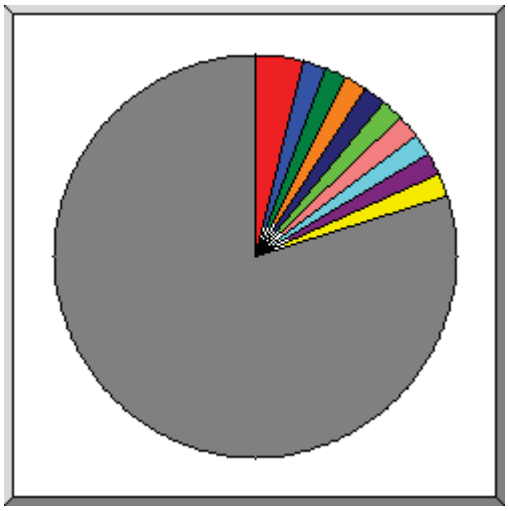
Dilimlerin temsil ettiği büyüklük: erişim miktarı.

/voleybol/
/yeniweb/
/cgi-bin/
/fakulte/
[Ana dizin]
/fotogaleri/
/yukokul/
/~mimarlik/
/images/
/~mnecat/
diğer

Dizinler listeleniyor - en az 0.01% trafiği olan, Sıralama: erişim miktarı.

erişim	bytes%	bölüm
20778	37.15%	/voleybol/
2315395	12.98%	/yeniweb/
140234	7.77%	/cgi-bin/
94814	7.09%	/fakulte/
192096	7.01%	[Ana dizin]
92760	3.60%	/fotogaleri/
89919	3.43%	/yukokul/
3378	2.91%	/~mimarlik/
932342	1.98%	/images/
21	1.93%	/~mnecat/
984	1.76%	/~eerdas/
14730	1.59%	/duyuru/
3782	1.28%	/~kimya/
5566	0.70%	/~yahyapamuk/
15552	0.61%	/dictur/
7875	0.59%	/~makina/
5864	0.58%	/~maden/
1837	0.57%	/~zoology/
2474	0.56%	/~baring/
7297	0.55%	/yokduyuru/
5317	0.43%	/ekart/
6104	0.41%	/enstitu/
122	0.35%	/~diclekus/
12905	0.23%	[not listed: 103 bölümler]

4.11. Erişim Raporu



Dilimlerin temsil ettiği büyüklük: erişim sayısı.

/
/yeniweb/images/images/anabar_01.jpg
/yeniweb/images/images/anabar_02.jpg
/yeniweb/images/images/anabar_03.jpg
/yeniweb/images/images/anabar_04.jpg
/yeniweb/images/bv01072.gif
/yeniweb/images/bv01077.gif
/yeniweb/images/bv01033.gif
/yeniweb/images/bv01034.gif
/yeniweb/images/bv01035.gif
diğer

dosyalar listeleniyor - en az 20 erişim, Sıralama: erişim sayısı.

erişim	bytes%	son saat	dosya
174672	6.22%	8/Ock/06 07:36	/
82363	0.31%	8/Ock/06 07:34	/yeniweb/images/images/anabar_01.jpg
81826	0.15%	8/Ock/06 07:34	/yeniweb/images/images/anabar_02.jpg
81295	0.03%	8/Ock/06 07:34	/yeniweb/images/images/anabar_03.jpg
81282	0.02%	8/Ock/06 07:34	/yeniweb/images/images/anabar_04.jpg
80660	0.03%	8/Ock/06 07:34	/yeniweb/images/bv01072.gif
80640	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01077.gif
80382	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01033.gif
80128	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01034.gif
79908	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01035.gif
79906	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01036.gif
79293	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01037.gif
78679	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01039.gif
77886	0.05%	8/Ock/06 07:34	/yeniweb/images/bv01080.gif
77144	0.09%	8/Ock/06 07:34	/yeniweb/images/bv01057.gif
76185	0.03%	8/Ock/06 07:34	/yeniweb/images/e-posta.gif

erişim	bytes%	son saat	dosya
72561	0.03%	8/Ock/06 07:34	/yeniweb/images/k1.gif
464		8/Ock/06 00:37	/yukokul/mmy/btzhsepd.gif
464	0.01%	8/Ock/06 02:48	/yeniweb/bilgiedinme/images/anabar.gif
463		8/Ock/06 00:37	/yukokul/mmy/blubul1a.gif
459		8/Ock/06 02:48	/yeniweb/bilgiedinme/images/ok.gif
458		8/Ock/06 00:37	/yukokul/mmy/button_anmo.gif
458	0.17%	8/Ock/06 06:11	/fotogaleri/fotogaleri/images/28.jpg
148		7/Ock/06 21:41	/fakulte/hukuk/1024x768
147		8/Ock/06 01:10	/~duhak/img/img/ukanat23.gif
20		8/Ock/06 01:25	/dictur/akademik/fakulte/katzir.html
71541	14.22%	8/Ock/06 07:36	[not listed: 18 998 dosyalar]

6. Sonuç

Dicle Üniversitesi sunucuları üzerinde tutulan access log dosyası 1-8 ocak 2006 kayıtları üzerinde yapılan incelemede aşağıdaki sonuçlar bulunmuştur

1. İnceleme yapılan dönem içerisinde; farklı IP lerden 463829 erişim olmuş ve 71GB bilgi transfer edilmiştir.
2. Sisteme en çok Çarşamba ve Perşembe günleri erişim olmuş ve en çok bu günlerde log oluşturmuştur.
3. Sistem kullanımı en çok 14 ve 15 saatlerinde oluşmuştur.
4. En çok erişim 85 ve 81 IP lerinden geldiği ve bu IP'lerin Telekomun ADSL kullanıcıları için ayırdığı IP bloğu olduğu ve bu IP lerden saldırı ve atakların geldiği tespit edilmiştir.
5. Gelen erişimlerin %50 den fazlası 85 li IP lerden olduğu ve bunun büyük bir oran olduğu görülmektedir. Sistemin en büyük kullanıcı grubunu oluşturan Kampus içinden kullanım yani 10 lu IP lerden gelen trafik bile %20 civarındadır. Bu da özellikle 85 li Ip lerden gelen istemin normal bir istem olmayıp ***kötü amaç*** taşıdığını göstermektedir.
6. Dosya istem kodları incelendiğinde en büyük oranın 200 OK kod olması gerekirken, 304 kod olduğu yani aynı dosyanın değişiklik olmadığı halde tekrar, tekrar istenmesi olduğu görülmekte, bu tekrar istemlerin de ***kötü amaç*** taşıdığı düşündürmektedir.
7. Dosya boyutu incelenmesinde en büyük oran %50 den fazlası 0 Kb görülmekte, bu da bize 6 madde ile uyumlu gelmektedir, çünkü sistem karşı tarafa bir kez gönderdiği dosyayı değişiklik yok ise tekrar göndermemektedir, aynı dosyanın tekrar, tekrar istemleri sistemi meşgul etmeye ve sistem kaynaklarını boşa harcamaya yöneliktir.
8. Dosya tiplerine bakıldığında en çok WMV (video) ve JPG dosya tiplerinin kullanıldığı, wmv nin ise Bayan Voleybol Takımının videosundan kaynaklandığı düşünülmektedir.
9. WMV dosya tipi 507 erişim ile trafiğin %35 ni, JPG dosya tipi ise 1668493 erişim ile trafiğin %25 oluşturmuştur, bu da bize WMV dosya tipine ulaşım az olmasına rağmen, dosya boyutlarının büyük olduğunu

göstermektedir. İlginç olan ise html/htm dosya tipinin %4 gibi çok düşük bir trafik oluşturması

10. Dizin ulaşımlarına bilgi transferi olarak bakıldığında 9. madde ile tutarlı olarak en büyük oranın /voleybol/ dizini olduğu görülmektedir. Ulaşım sayısı olarak bakıldığında ise /yeniweb/ olduğu görülmektedir
11. Sonuç olarak iyi niyetli olmayan atakların oluştuğu gözlemlenmiştir

Kaynakça

1. Takıçı H., Soğukpınar H.: "Saldırı Tespitinde Yeni bir Yaklaşım" , Gebze Yüksek Teknoloji Enstitüsü, http://www.bilmuh.gyte.edu.tr/datamining/ files/b2002-saldiri-tesbiti-h-takci.doc
2. Etzioni, O.: (1996). "The World-Wide Web: Quagmire or Gold Mine?"
3. Cooley, Robert., Mobasher, Bamshad., Srivastava, Jaideep: (1997). "Web Mining: Information and Pattern Discovery on the World Wide Web"
4. Garofalakis, Minos N., Rastogi, Rajeev. Bell Laboratories (1999). "Data Mining and the Web: Past, Present and Future"
5. Dayıoğlu B.:, "Elektronik Saldırı Tespiti", http://www.teknoturk.org/docking/yazilar/tt000026-yazi.htm
6. P Joshi, Karuna., Joshi, Anupam.: (1999). "Warehousing and Mining Web Logs"